Tijdschrift voor Medisch Onderwijs, maart 2005 | Vol. 24, nr. 1, p. 14-23

De ontwikkeling van een feedbacksysteem voor toetsvragenmakers

J.J. Reinders, J. Cohen-Schotanus, W.M. Molenaar

Samenvatting

Door de Groningse Faculteit der Medische Wetenschappen is een feedbacksysteem voor toetsvragenmakers ontwikkeld. Het systeem is onder andere gebaseerd op een statistische analyse van de toetsresultaten. Uit een eerste peiling onder toetsvragenmakers blijken de respondenten overwegend positief te zijn over het ontvangen van feedback. Het ontwikkelde feedbacksysteem is echter nog voor verbetering vatbaar. Om het te kunnen optimaliseren, worden de suggesties van de respondenten als handvat gebruikt. Het specificeren van de feedbackinformatie en het verhelderen van de procedures rond de eliminatie van toetsvragen zijn daarbij twee van de belangrijkste aandachtspunten. (Reinders JJ, Cohen-Schotanus J, Molenaar WM. De ontwikkeling van een feedbacksysteem voor toetsvragenmakers. Tijdschrift voor Medisch Onderwijs 2005;24(1):14-23.)

Inleiding

Binnen het onderwijs laten studenten zich in de eerste plaats leiden door de eisen met betrekking tot hun examens. Studenten bestuderen vooral datgene waar zij op worden beoordeeld.[1-4] Mede daarom is de kwaliteit van een toetsingsprogramma van groot belang, en staat het verbeteren van de toetsing hoog op de agenda van de Faculteit der Medische Wetenschappen (FMW) te Groningen.[5] Toetsen die slecht van kwaliteit zijn, meten niet wat ze zouden moeten meten. Wenselijke studentprestaties kunnen dan ook niet goed worden geëvalueerd en op rechtvaardige wijze worden beloond.[6-7]

Toetsing binnen de FMW bestaat voor een groot deel uit schriftelijke toetsen met geprecodeerde toetsvragen. De kwaliteit van deze laat echter vaak te wensen over.[8] Ze bevatten onder andere dubbelzinnige termen, semi-kwantitatieve beweringen, dubbele ontkenningen en normatieve uitspraken. Teneinde de kwaliteit te verbeteren heeft de FMW een traject ingezet om docenten te ondersteunen bij het construeren van toetsvragen.

Allereerst is een intervisie-overleg voor toetsvragenmakers in het leven geroepen (de Toetsservice), waarbij 'toetscoaches' toetsvragenmakers begeleiden bij het verbeteren van toetsvragen.[9] Deze Toetsservice vindt plaats voordat de toets bij de student wordt afgenomen. Vervolgens is een feedbacksysteem voor toetsvragenmakers ontwikkeld dat is gebaseerd op een statistische analyse van de toetsresultaten. De feedback aan toetsvragenmakers moet docenten meer inzicht geven in de kwaliteit van hun toetsvragen. Daarbij moet de feedback aanknopingspunten geven om de kwaliteit van de toetsvragen systematisch te verbeteren. Daarnaast zullen toetsvragenmakers op hun verantwoordelijkheden moeten worden aangesproken, zodat sturing kan worden gegeven aan hun werkprestaties en vrijblijvendheid bij het construeren van toetsvragen kan worden verminderd. De feedback kan daarbij formatieve en summatieve doelen hebben.[10] Summatieve feedback is gericht op het beoordelen van bekwaamheid en wordt gebruikt op het moment dat er beslissingen worden genomen over de be-

trokken toetsvragenmaker (bijv. certificeren). Het gaat hier om de vergelijking van individuele competenties met een algemene kwalificatiestandaard. Formatieve feedback is gericht op de ontwikkeling van competenties en wordt gebruikt om vorderingen in beeld te brengen, te reflecteren op het leerproces, conclusies te trekken en plannen te maken voor verdere competentieontwikkeling. Formatieve feedback ondersteunt het leer- en ontwikkelproces. Het verschil tussen summatief en formatief evalueren of feedback geven ligt met andere woorden niet in de manier waarop men beoordeelt, maar wel in wat men doet met deze beoordeling. Het is, naast het leerproces van de docent, vooral belangrijk de docent aan te spreken op zijn verantwoordelijkheden. Daarom is het de bedoeling dat de feedback aan toetsvragenmaker op termijn een summatief karakter krijgt.

Dit artikel beschrijft de ontwikkeling van het feedbacksysteem voor toetsvragenmakers en de resultaten van een eerste peiling onder toetsvragenmakers die feedback ontvingen.

De ontwikkeling van het feedbacksysteem

Teneinde de ontwikkeling van een feedbacksysteem richting te kunnen geven en te kunnen evalueren, zijn een aantal voorwaarden geformuleerd waar het systeem aan moet voldoen. Het feedbacksysteem moet *efficiënt en gebruiksvriendelijk* zijn en *bruikbaar* voor de feedbackontvanger. De bruikbaarheid van de feedback voor de feedbackontvanger is gedefinieerd in termen van *concreetheid, veiligheid* en *toepasbaarheid*.[11] Ook moet de toetsvragenmaker verantwoordelijk worden gesteld voor de kwaliteit van zijn/haar werk. Het feedbacksysteem zal daarom tevens mogelijkheden moeten bieden om summatief feedback te kunnen geven. Het feedbacksysteem zal dan dienst doen als beoordelingsinstrument. Daarnaast moeten de persoonlijke feedbackgegevens *vertrouwelijk* worden verwerkt in zoverre dat rekening wordt gehouden met de privacy van de docent. Tot slot moet de feedbackprocedure praktisch en technisch *haalbaar* zijn.

Uit literatuuronderzoek naar feedbacksystemen voor docenten blijken meerdere soorten feedbacksystemen te bestaan.[12-13] Op de kwaliteit van toetsvragen wordt echter weinig feedback gegeven. Systemen die wel voor feedback aan toetsvragenmakers gebruikt zouden kunnen worden, zijn in de eerste plaats bestemd voor digitale toetsing (o.a. Questionmark Perception en Egel) of zijn onderdeel van een elektronische leeromgeving (o.a. Blackboard). Veel systemen zijn echter vooral bestemd voor de constructie en opslag van toetsvragen.[15] De mogelijkheden van de laatstgenoemde systemen beperken zich meestal tot simpele statistische analyses en geven te weinig aanwijzingen waarom een toetsvraag slecht van kwaliteit zou zijn. Bovendien is digitale toetsing binnen de FMW niet gebruikelijk. Om die reden kunnen dit soort feedbackmogelijkheden niet worden benut.

Gegeven de bovengenoemde bevindingen is besloten een 'eigen' feedbacksysteem te ontwikkelen. Van belang zijn daarbij de potentiële feedbackbronnen, de te ontwikkelen feedbacksoftware en de feedbackrapportage. Tot slot zijn de technische en procedurele aspecten van het feedbacksysteem uitgewerkt. Het ontwikkeltraject heeft geresulteerd in een feedbacksysteem dat bestaat uit een softwarepakket en een aantal standaardprocedures. De gekozen feedbackbronnen met de gekozen kwaliteitscriteria, de feedbacksoftware en het ontwerp van de feedbackrapportage zullen hieronder nader worden beschreven.

Feedbackbronnen

De itemanalyse vormt de eerste feedbackbron van het feedbacksysteem. Twee indices van de itemanalyse blijken het meest geschikt als criteria om uitspraken te doen over de kwaliteit van individuele toetsvragen: de P-waarde en de Rit-waarde.[16-17] De P-waarde, ook wel moeilijkheidsgraad genoemd, geeft de proportie van examinandi die een bepaalde vraag (item) correct hebben beantwoord. De range van deze index ligt tussen de 0 (niemand heeft de vraag goed) en 1 (iedereen heeft de vraag goed). De Rit of item-totaal correlatie geeft het discriminerend vermogen van een item aan. De range van deze index ligt tussen de –1 en +1. Hoe hoger de Rit, hoe meer onderscheid wordt gemaakt tussen goed en slecht presterende studenten. Een negatieve Rit betekent dat vooral de slecht presterende studenten het item juist beantwoorden. Met betrekking tot de interpretatie van bovengenoemde criteria blijken verschillende kwaliteitsnormen te bestaan. Afhankelijk van de onderwijskundige uitgangspunten zijn meerdere statistische kwaliteitsnormen acceptabel.[17-20] Deze onderwijskundige uitgangspunten hebben met name betrekking op de functie van de toets. Een toets is selecterend als hij alleen bij voldoende resultaat toegang geeft tot een volgend studieonderdeel of de volgende periode.[21] De selecterende functie van de toets uit zich in het discriminerend vermogen van de afzonderlijke vragen. In dat geval zal de Rit-waarde per vraag tenminste positief en significant moeten zijn. Daarbij is er, naast vooraf vastgestelde normen, ook een toetsafhankelijke norm. De itemanalyse geeft per toets het niveau aan waarbij de Rit significant is. De groepsomvang (aantal examinandi) heeft invloed op de laatstgenoemde Rit-norm. Een standaard Rit-norm zal onwenselijk zijn, omdat de groepsomvang per opleiding binnen de FMW sterk uiteenloopt. Wanneer men een toets wil hebben die gedifferentieerd is wat betreft de opbouw in niveau van toetsvragen binnen een toets, dan kan men toetsvragen met een P-waarde groter of gelijk aan .90 accepteren. Ook minimaal vereiste kennis wordt op deze wijze getoetst. Wanneer de P-waarde groter is dan .90 zal een toetsvraag echter niet goed meer in staat zijn te discrimineren tussen goede en slechte studenten en zal de Rit overwegend laag en weinig betekenisvol zijn. Om die reden is gekozen om bij toetsvragen met een P-waarde groter of gelijk aan .90 een Rit-waarde toe te laten die in ieder geval niet significant negatief is. Wanneer een toetsvraag een lagere P-waarde heeft dan de giskans, hebben minder studenten het item goed dan op basis van kans verwacht kan worden. Zulke toetsvragen zijn altijd 'verdacht'.[17]

De tweede feedbackbron heeft betrekking op de nabespreking van de toetscoördinator met de Jaarvertegenwoordiging (studentafgevaardigden die de tentamenevaluatie verwerken). Naar aanleiding van die nabespreking kunnen door de toetscoördinator toetsvragen worden geëlimineerd en verkeerde sleutels worden gewijzigd. In de eerste plaats zijn vraageliminatie en sleutelwijzigingen bedoeld om tot een rechtvaardige beoordeling te komen. Onjuist geformuleerde vragen of incorrecte antwoordsleutels beïnvloeden immers slagen en zakken voor een toets. Ten tweede behoort een toets tenminste betrouwbaar te zijn. Door vragen te elimineren en incorrecte sleutels te wijzigingen, kan de toetscoördinator de betrouwbaarheid van de gehele toets verbeteren.[22] Wanneer een toetsvraag niet aan de gewenste normen voldoet, dan zal de toetscoördinator de toetsvraag nader kunnen onderzoeken. Het studentcommentaar uit de tentamenevaluatie en het oordeel van één of meer inhoudsdeskundigen kunnen de toetscoördinator ondersteunen in zijn

beslissing vragen te elimineren of incorrecte antwoordsleutels op te sporen en te wijzigen. De toetsvragenmaker kan via de feedback achterhalen of zijn toetsvragen zijn geëlimineerd en of zijn antwoordsleutels al dan niet correct waren. Vijf categorieën worden gehanteerd om de reden van eliminatie aan te geven: 1) psychometrisch/toets-statistisch, 2) vormtechnisch, 3) inhoudelijk, 4) traceerbaar in de leerstof en 5) overig.

Samenvattend zijn uit bovengenoemde overwegingen de volgende kwaliteitsnormen vastgesteld.Een toetsvraag voldoet aan de statistische normen als in de itemanalyse de P-waarde groter is dan de giskans, de Rit tenminste positief significant is en, wanneer de P-waarde groter dan of gelijk aan .90 is, de Rit tenminste niet negatief significant is. Op basis van onwenselijke P- en Rit-waarden, studentcommentaar en het oordeel van een inhoudsdeskundige kunnen toetsvragen worden geëlimineerd en incorrecte antwoordsleutels worden opgespoord.

Feedbacksoftware

Het feedbacksoftwarepakket bestaat uit een aantal Access-databases die gekoppeld zijn aan een applicatie. Outputbestanden van statistische analyses worden ingelezen en vervolgens opgeslagen in één van de databases. Daarnaast kan aanvullende informatie handmatig worden ingevoerd via keuzemenu's. Ook namen en adresgegevens van toetsvragenmakers zijn in een database opgeslagen waarbij het systeem de mogelijkheid geeft gegevens te wijzigen of aan te vullen. Om eventuele invoerfouten te voorkomen laat het systeem meldingen zien wanneer verkeerde operaties worden uitgevoerd.

De ontwikkelde feedbackapplicatie heeft twee soorten output. Ten eerste maakt zij een feedbackgrafiek waarbij de P-waarden en de Rit-waarden van de vragen tegen elkaar worden uitgezet. Elke toetsvragenmaker ontvangt een overzicht van alle toetsvragen, waarbij van zijn eigen vragen de vraagnummers zichtbaar zijn. Vragen van andere toetsvragenmakers zijn niet als nummer zichtbaar, maar worden geanonimiseerd en door zwarte punten gerepresenteerd. Ten tweede maakt de applicatie een soort kengetallenpagina aan. Hierop staan algemene toetsgegevens, redenen van vraageliminatie en sleutelwijzigingen.

Feedbackrapportage

Toetsvragenmakers ontvangen per toets een pakket bestaande uit: 1) een kengetallenpagina (figuur 1), 2) een feedbackgrafiek inclusief legenda (figuur 2) en 3) een versie van de toets zoals deze uiteindelijk is afgenomen. Omdat elk type geprecodeerde toetsvraag (de twee-, drie- of vierkeuze vraag) een eigen giskans heeft, is omwille van de inzichtelijkheid gekozen een aparte grafiek te maken voor elk type geprecodeerde vraag. Door verschillende giskansen zullen de statistische normen per type geprecodeerde toetsvraag immers verschillen.

Feedback aan toetsvragenmakers: een peiling onder docenten

Het feedbacksysteem is geëvalueerd door middel van een enquête. De aangeschreven groep bestond uit 54 docenten van de opleidingen Tandheelkunde en Geneeskunde van de Faculteit der Medische Wetenschappen van de Rijksuniversiteit Groningen.

Aan respondenten werd gevraagd 4 gesloten vragen en één open vraag te beantwoorden die betrekking hebben op de feedbackrapportage: 1) "Ik vind de door mij ontvangen grafiek(en) inzichtelijk", 2) "Ik vind de toelichting bij de door mij ontvangen grafiek(en) begrijpelijk", 3) "De door mij ontvangen feedback kan ik goed gebruiken om de volgende keer mijn

Toetsevaluatie
Deze feedback heeft betrekking op: driekeuze vragen

Persoonsgegevens
Alle Toetsvragenmakers

Algemene gegevens

Studie:	Geneeskunde	Aantal 2-keuze vragen:	85
Toets:	GNK X.X	Aantal 3-keuze vragen:	30
Datum toetsafname:	XX-XX-XXXX	Aantal 4-keuze vragen:	14
		Aantal examinandi:	278
Percentage geslaagden:	87 %		
RIT-grenswaarde:	0,12		

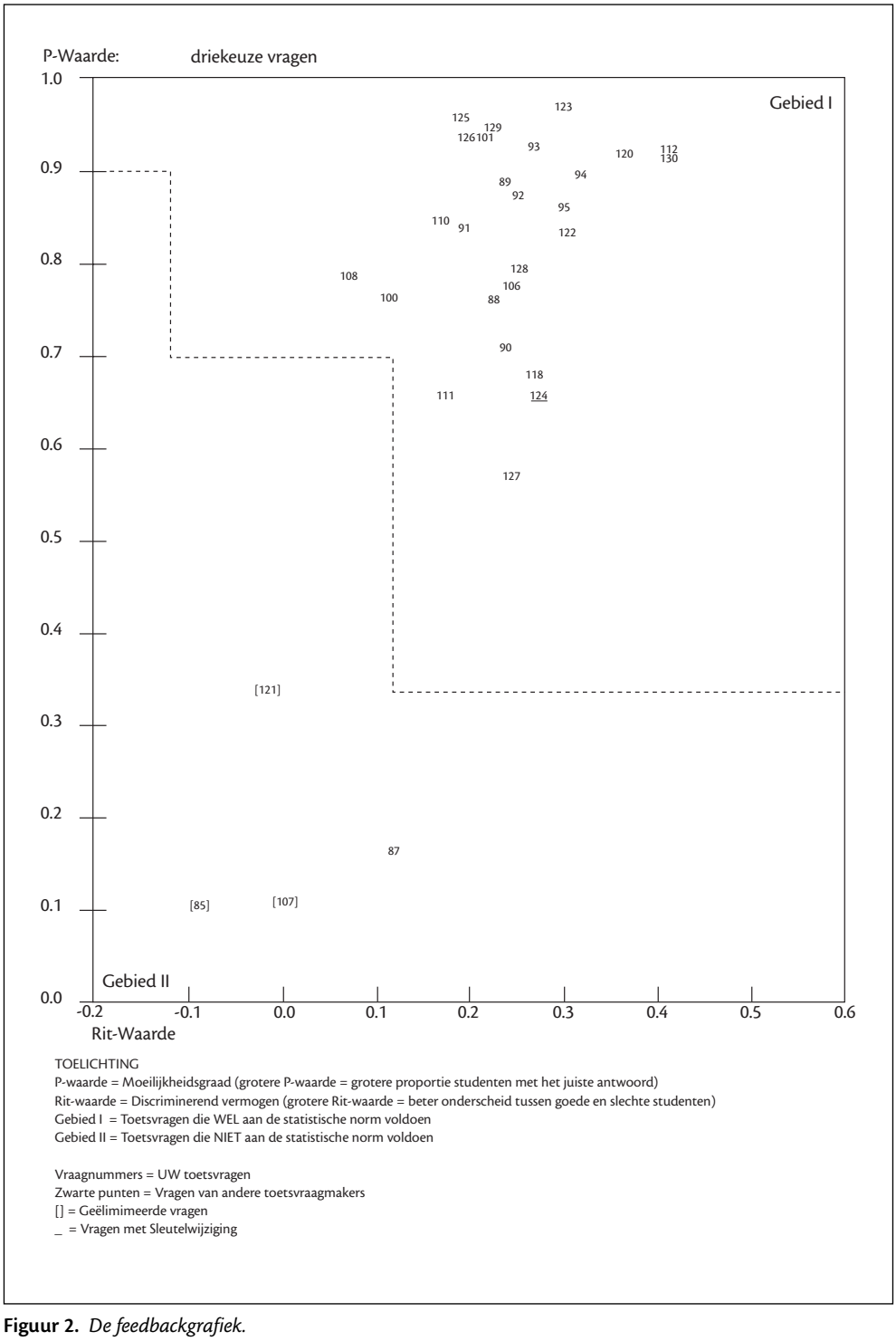
Type toetsvraag	**Gemiddelde P-waarde**	**Toetsvragen aangeleverd** (aantal vragen)	**Voldoet NIET aan statistische normen** (aantal vragen)	**Toetsvragen geëlimineerd** (aantal vragen)
Alle 3-keuze vragen:	0,74	30	6	3

Reden van eliminatie	**Geëlimineerde vragen** (vraagnummers)
Psychometrische/ toetsstatistische redenen:	86, 107, 121
Vormtechnische redenen:	86
Inhoudelijke redenen:	107, 121
Ontraceerbaar in leerstof:	-
Overige redenen:	-
Vragen met sleutelwijziging: (vraagnummers)	124

Feedbackgrafiek toetsananalyse: zie bijlage

Faculteit Medische Wetenschappen RUG
Onderwijsinstituut - Ontwikkeling en Kwaliteitszorg

Figuur 1. *De kengetallenpagina.*



Figuur 2. *De feedbackgrafiek.*

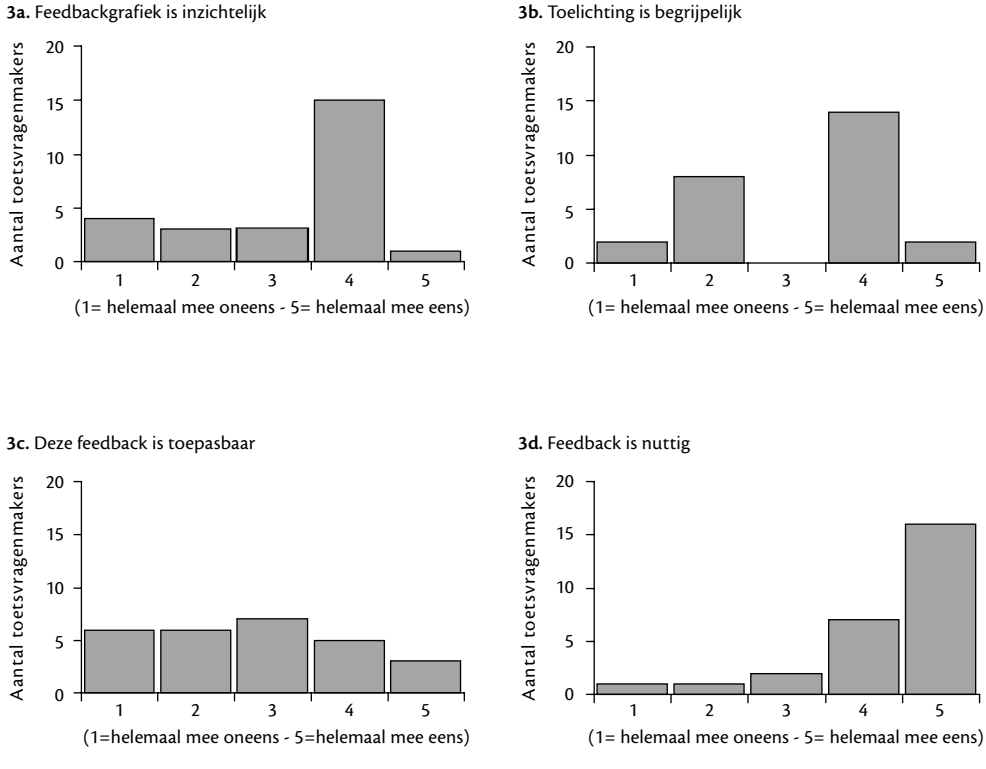
toetsvragen te verbeteren", 4) "Ik vind het geven van feedback aan toetsvragenmakers een goed idee". De gesloten vragen werden beantwoord op een vijf-punts schaal (1= helemaal mee oneens tot 5= helemaal mee eens). De open vraag betrof: 5) "Wat zou, volgens u, aan deze feedback moeten worden veranderd?" Enquête en feedbackrapportage werden tegelijkertijd door de aangeschreven groep ontvangen.

De respons op de enquête was 54% (n=29). Het merendeel van de respondenten is tevreden over de inzichtelijkheid van de grafiek (n=16; figuur 3a) en de begrijpelijkheid van de toelichting bij de grafiek (n=16; figuur 3b). De meningen over de toepasbaarheid van de feedback zijn verdeeld (figuur 3c). Een overgrote meerderheid vindt het geven van feedback aan toetsvragenmakers nuttig (n=23; figuur 3d).

Voorts is de respondenten gevraagd welke suggesties zij hebben voor verbetering van het feedbacksysteem (tabel 1). Vijftien opmerkingen hadden betrekking op de behoefte aan meer specifieke informatie. Twintig opmerkingen betroffen onduidelijkheden rond de eliminatie van toetsvragen en de keuze van statistische normen. Overige opmerkingen hadden met name betrekking op feedbackprocedures en motivationele effecten van deze vorm van feedback.

Discussie

Een feedbacksysteem voor toetsvragenmakers is ontwikkeld, waarbij een statisti-



Figuur 3. *Enquête toetsvragenmakers.*

Tabel 1. *Opmerkingen van toetsvragenmakers/feedbackontvangers.*

Aard opmerking	Beschrijving opmerking	Frequentie
Aard feedbackinformatie	Graag ook studentcommentaar	8
	Graag ook commentaar toetscoördinator	3
	Graag ook nagesprek	3
	Graag meer feedbackinformatie	1
	Totaal	15
Toelichting	Statistische normen onduidelijk	6
	Eliminatieprocedure onduidelijk of onbekend	9
	Toelichting graag uitgebreider	5
	Totaal	20
Overig	Ondersteuning voor toetsvragenmakers is onbekend	2
	Selectie van eigen vragen i.p.v. totale toets op papier (papierverspilling)	2
	Eerst uitdraai grafiek voor toetscoördinator, tweede uitdraai voor feedback	1
	Enkele eigen vragen ontbreken (informatie van coördinator incompleet of incorrect)	4
	Deze feedback is contra-effectief/leidt tot demotivatie	5
	Graag tijdig ontvangst van feedback	1
	Docent betrekken bij eliminatieprocedure	1

sche analyse van de toetsresultaten de belangrijkste feedbackbron vormt. Het ontwikkelde feedbacksysteem voldoet echter nog niet aan alle uitgangspunten die bij de start van het ontwikkelingsproject zijn geformuleerd. Hoewel de respondenten overwegend positief zijn over het ontvangen van feedback verdienen de efficiëntie en de bruikbaarheid van het feedbacksysteem meer aandacht. De bruikbaarheid van de feedback is te verbeteren door meer en concretere informatie te verstrekken. Daarbij moet de feedbackontvanger beter worden ingelicht over de eliminatieprocedure zoals deze binnen de FMW wordt gehanteerd.

De respons op de enquête was relatief laag. Mogelijk bestaat de non-respons uit docenten die geen commentaar hadden op de feedback. Daarnaast zijn er enkele oningevulde vragenlijsten teruggestuurd die echter wel werden begeleid met schriftelijk commentaar. Dit laatstgenoemde commentaar had betrekking op de billijkheid van de feedback en houdt zeer waarschijnlijk verband met de te bondige toelichting bij de feedbackrapportage.

Een kanttekening is dat het feedbacksysteem niet voor elke situatie geschikt is. Alleen over toetsen met meer dan 50 examinandi kan feedback worden gegeven. De P-waarde is, volgens Van Os, bij een minimum van 25 examinandi voldoende stabiel om uitspraken te doen over de moeilijkheidsgraad.[19] Bij minder dan 25 personen verliest de P-waarde aan betekenis. Wanneer het aantal deelnemers aan een toets 50 of minder is, zal de Rit sterk fluctueren en dus niet langer betrouwbaar zijn. De waarden van deze index worden dan extreem hoog of juist extreem laag.[19] Dit kan vooral voor de opleiding Tand-

heelkunde tot problemen leiden vanwege haar kleine studentenaantal. Daarnaast is het feedbacksysteem niet geschikt voor hertentamens. De 'goede' student zal door de reguliere toets immers zijn uitgeselecteerd en doorgaans wordt een hertentamen bij een kleine groep examinandi afgenomen.

Om het feedbacksysteem te kunnen optimaliseren, worden de suggesties van de respondenten als handvat gebruikt. Het specificeren van de feedbackinformatie en het verhelderen van de procedures rond de eliminatie van toetsvragen zijn daarbij twee van de belangrijkste aandachtspunten. Inmiddels is binnen de FMW een standaard eliminatieprocedure vastgesteld. De procedure bestaat uit een aantal statistische normen. Aan de hand van deze normen zijn drie condities mogelijk: 1) een vraag voldoet aan de normen en blijft onderdeel van de toets, 2) een vraag voldoet niet aan de normen en wordt direct geëlimineerd en 3) een vraag is van twijfelachtige kwaliteit en dient nader te worden onderzocht. In het laatste geval kan nadere bestudering van een mogelijke 'probleemvraag' leiden tot het besluit de vraag te elimineren. Toekomstig onderzoek zal moeten uitwijzen of deze feedback aan toetsvragenmakers daadwerkelijk een rol speelt bij de verbetering van de kwaliteit van geprecodeerde toetsen. Om het feedbacksysteem in de toekomst bruikbaar te maken voor summatieve doeleinden, is het noodzakelijk om helder en eenduidig beleid te ontwikkelen rond de beoordeling en kwalificatie van docenten en toetsvragenmakers in het bijzonder.

Literatuur

1. Eringa K, Rietveld J, Zwaal W. Assessment en assessmentcenters. De leerwaarde van toetsen. In: Heijnen GW, Meeder S, editors. Toetsen en ICT in het hoger onderwijs. Utrecht: Stichting SURF; 2000.
2. Tait B. Constructive internet based learning [homepage op internet]. Londen: University of East London; c1997-2005 [geciteerd 14 januari 2005]. Beschikbaar op: http://homepages.uel.ac.uk/W.H.Tait/Learning/Cibl.htm.
3. Vleuten CPM van der. Toetsing van medische competentie. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der. Medisch onderwijs in de praktijk. Assen: Van Gorcum; 1995.
4. Cohen-Schotanus J. Student assessment and examination rules. Med Teach 1999;21(3):318-21.
5. Cohen-Schotanus J. Kwaliteit van toetsen. Beleidsnotitie voor toetsprogramma's opleidingen FMW RuG. Groningen: Faculteit der Medische Wetenschappen, Rijksuniversiteit Groningen; juli 2003.
6. Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Med Educ 1998; 32:343-50.
7. Downing SM. Construct-irrelevant variance and flawed test questions: do multiple-choice item-writing principles make any difference? Acad Med 2002 oktober;77(10 Suppl):S103-S104.
8. Lang G. Multiple Choice vragen zijn uit de tijd. Open brief in UK. Groningen: Rijksuniversiteit Groningen. 28 februari 2002;nr. 24.
9. Dollekamp BA. Verslag toetsservice FMW. Periode maart 2002-augustus 2003. Groningen: Afdeling OWI-OK, Werkgroep Kwalificatie en scholing van docenten, Faculteit der Medische Wetenschappen, Rijksuniversiteit Groningen; oktober 2003.
10. Kolk J van der, Willemse P. Van eindtermen naar competentiebeoordeling. Vernieuwing van de examinering in de BVE-sector. 's-Hertogenbosch: CINOP; 2002.
11. Spaendonck KPM van, Weel-Baumgarten EM van. Feedback. In: Metz JCM, Scherpbier AJJA, Houtkoop E, editors. Gezond Onderwijs-II 1993;2:62-5.
12. Odden A. Cincinnati's new approach to teacher compensation. School Business Affairs 2002;68(5):20-3.
13. Stronge JH, Tucker PD. Handbook of teacher evaluation: assessing and improving performance. New York: Eye On Education; 2003.
14. Feenstra H. Beoordelen en technologie. In: Berkel H van, Bax A. Toetsen in het hoger onderwijs. Houten/Diegem: Bohn Stafleu Van Loghum; 2002. p. 121-2.
15. Dousma T, Hortsen A, Brants J. Tentamineren. Groningen: Wolters-Noordhoff; 1997. p. 126-8,132-5.
16. Bender W. Item-analyse. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, editors. Medisch onderwijs in de praktijk. Assen: Van Gorcum; 1995.
17. Ebel RL. Essentials of educational measurement. New York: Prentice Hall; 1972.

18. Os W van. Evaluatie in het hoger onderwijs: controle op en verbetering van de kwaliteit van hoger onderwijs. Groningen: Wolters-Noordhoff; 1987.
19. Groot AD de, Naerssen RF van. Studietoetsen construeren, afnemen, analyseren. Den Haag: Mouton; 1969.
20. Pouw C, Boshuizen E. Onderwijsontwikkeling voor kwalificatieontwikkeling. In: Dam A ten, et al, editors. Onderwijskunde hoger onderwijs. Handboek voor docenten. Assen: Van Gorcum; 2000.
21. Hogeschool van Amsterdam. Kwaliteitsverbetering met MC toetsvragen [homepage op internet]. Amsterdam: Hogeschool van Amsterdam; c2002-2005 [geciteerd 14 januari 2005]. Beschikbaar op: http://www.oro.hva.nl/detailscherm.php?obj_id=1349.

De auteurs:
Drs. J.J. Reinders is projectmedewerker.
Mw. dr. J. Cohen-Schotanus is afdelingshoofd.
Mw. prof. dr. W.M. Molenaar is voorzitter van de Werkgroep Kwalificatie en Scholing van Docenten.
Allen zijn verbonden aan de afdeling Onderwijsontwikkeling en Kwaliteitszorg van het Onderwijsinstituut, Faculteit der Medische Wetenschappen, Rijksuniversiteit Groningen.

Correspondentieadres:
Drs. J.J. Reinders, Onderwijsinstituut, Afdeling Onderwijsontwikkeling en Kwaliteitszorg, Faculteit der Medische Wetenschappen, Rijksuniversiteit Groningen, A. Deusinglaan 1, 9713 AV Groningen, tel.: 050-3637349, fax: 050-3633865, j.j.reinders@med.rug.nl.

Summary

A system of providing feedback to authors of test items was developed by the Faculty of Medical Sciences, University of Groningen, the Netherlands. The feedback system is partly based on statistical analyses of multiple choice tests. An initial survey among authors of test items revealed that authors were generally willing to receive feedback. The feedback system that was developed can be improved, however. We will use item authors' suggestions as starting points for measures to optimise the feedback system. These measures will be focused on the provision of more detailed feedback and clarification of the procedures that are used in decisions about the elimination of test items. (Reinders JJ, Cohen-Schotanus J, Molenaar WM. Developing a system of providing feedback to authors of test items. Dutch Journal of Medical Education 2005;24(1):14-23.)